Bounds for left and right window cutoffs

Javiera Barrera, Bernard Ycart

Bounds for left and right window cutoffs

Dedicated to the memory of Béatrice Lachaud

Javiera Barrera* Bernard Ycart†

October 2, 2013

Abstract

The location and width of the time window in which a sequence of processes converges to equilibrum are given under conditions of exponential convergence. The location depends on the side: the left-window and right-window cutoffs may have different locations. Bounds on the distance to equilibrium are given for both sides. Examples prove that the bounds are tight.
Keywords: cutoff; exponential ergodicity
MSC: 60J25

1 Introduction

The term "cutoff" was introduced by Aldous and Diaconis [1], to describe the phenomenon of abrupt convergence of shuffling Markov chains. Many families of stochastic processes have since been shown to have similar properties: see [13, Chap. 8] for an introduction to the subject, [16] for a review of random walk models in which the phenomenon occurs, and [4] for an overview of the theory. Consider a sequence of stochastic processes in continuous time, each converging to a stationary distribution. Denote by $d_n(t)$ the distance between the distribution at time t of the n-th process and its stationary distribution, the 'distance' having one of the usual definitions (total variation, separation, Hellinger, relative entropy, L^p, etc.). The phenomenon can be expressed at three increasingly sharp levels (more precise definitions will be given in section 2).

1. The sequence has a cutoff at (t_n) if $d_n(ct_n)$ tends to the maximum M of the distance if $c < 1$, to 0 if $c > 1$.
2. The sequence has a window cutoff at (t_n, w_n) if $\liminf d_n(t_n + cw_n)$ tends to M as c tends to $-\infty$, and $\limsup d_n(t_n + cw_n)$ tends to 0 as c tends to $+\infty$.
3. The sequence has a profile cutoff at (t_n, w_n) with profile F if $F(c) = \lim d_n(t_n+cw_n)$ exists for all c, and F tends to M at $-\infty$, to 0 at $+\infty$.

There are essentially two ways to interpret the cutoff time t_n: as a mixing time [13, Chap. 18], or as a hitting time [14]. For samples of Markov chains, the latter interpretation can be used to determine explicit online stopping times for MCMC algorithms [18, 11, 12, 9].

Sequences of processes for which an explicit profile can be determined are scarce. The first example of a window cutoff concerned the random walk on the hypercube for the

*Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez. Av. Diagonal las Torres 2640 Peñalolén, Santiago, Chile. javiera.barrera@uai.cl

†Laboratoire Jean Kuntzmann, Univ. Grenoble-Alpes, 51 rue des Mathématiques 38041 Grenoble cedex 9, France. Bernard.Ycart@imag.fr

total variation distance; it was treated by Diaconis and Shahshahani shortly after the introduction of the notion [8]. It was soon precised into a profile cutoff by Diaconis, Graham, and Morrison [6]. Cutoffs for random walks on more general products or sums of graphs have been investigated in [19], and more recently by Miller and Peres [15]. Random walks on the hypercube can be interpreted as samples of binary Markov chains. Diaconis et al.'s results were generalized to samples of continuous and discrete time finite state Markov chains for the chi-squared and total variation distance in [17], then to samples of more general processes, for four different distances in [2, section 5] (see also [13, Chap. 20]). Other examples of profile cutoffs include the riffle shuffle for the total variation distance [3], and birth and death chains for the separation distance [7] or the total variation distance [10]. When the maximum M of the distance is 1 (total variation, separation), the profile F decreases from 1 to 0. Thus it can be seen as the survival function of some probability distribution on the real line. A Gaussian distribution has been found for the riffle shuffle with the total variation distance [3, Theorem 2] or for some birth and death chain with the separation distance [7, Theorem 6.1]. A Gumbel distribution has been found for samples of finite Markov chains and the total variation distance [6, 17]. For the Hellinger, chi-squared, or relative entropy distances, other profiles were obtained in [2].

Explicit profiles are usually out of reach, in particular for the total variation distance: only a window cutoff can be hoped for. However the definition above, which is usually agreed upon ([4, Definition 2.1] or [13, p. 218]), may not capture the variety of all possible situations. As will be shown here, the location of a left-window cutoff should be distinguished from that of a right-window cutoff: see Figure 18.2, p. 256 of [13]. The main result of this note, Theorem 2.1, expresses the characteristics of the left and right windows in terms of a decomposition into exponentials of the distances $d_n(t)$. It refines some of the results in Chen and Saloff-Coste [5], in particular Theorem 3.8. Explicit bounds on the distance to equilibrium are given. They are proved to be tight, using examples of cutoffs for Ornstein-Uhlenbeck processes (see Lachaud [11]).

The paper is organized as follows. Section 2 contains formal definitions and statements. Examples are given in section 3. Theorem 2.1 is proved in section 4.

2 Definitions and statements

For each positive integer n a stochastic process $X_n = \{X_n(t)\,;\ t \geqslant 0\}$ is given. We assume that $X_n(t)$ converges in distribution to ν_n as t tends to infinity. The convergence is measured by one of the usual distances (total variation, separation, Hellinger, relative entropy, L^p, etc.), the maximum of which is denoted by M ($M = 1$ for total variation and separation, $M = +\infty$ for relative entropy, chi-squared...). The distance between the distribution of $X_n(t)$ and ν_n is denoted by $d_n(t)$.

Definition 2.1. *Denote by (t_n) and (w_n) two sequences of positive reals, such that $w_n = o(t_n)$. They will be referred to as respectively as* location *and* width. *The sequence (X_n) has:*

1. *a left-window cutoff at (t_n, w_n) if:*
$$\lim_{c\to-\infty} \liminf_{n\to\infty} \inf_{t<t_n+cw_n} d_n(t) = M\ ;$$
2. *a right-window cutoff at (t_n, w_n) if:*
$$\lim_{c\to+\infty} \limsup_{n\to\infty} \sup_{t>t_n+cw_n} d_n(t) = 0\ ;$$
3. *a profile cutoff at (t_n, w_n) with profile F if:*
$$\forall c \in \mathbb{R}\,,\ F(c) = \lim_{n\to\infty} d_n(t_n + cw_n)$$

exists and satisfies:

$$\forall c \in \mathbb{R}\,,\ 0 < F(c) < M \quad \textit{and} \quad \lim_{c\to-\infty} F(c) = M\ , \quad \lim_{c\to+\infty} F(c) = 0\ .$$

If both left- and right-window cutoffs hold for the same location t_n and width w_n, then a (t_n, w_n)-cutoff holds in the sense of Definition 2.1 in Chen and Saloff-Coste [4]. The location and width are not uniquely determined. Observe that if a left-window cutoff holds at location t_n, it also holds at any location t'_n such that $t'_n \leqslant t_n$. Symmetrically, if a right-window cutoff holds at location t_n, it also holds at any location t'_n such that $t'_n \geqslant t_n$. Moreover, if a cutoff holds for width w_n, it also holds for any width w'_n such that $w'_n \geqslant w_n$. The location and width of a left-window cutoff will be said to be optimal if for any $c < 0$:

$$\liminf_{n\to\infty} \inf_{t<t_n+cw_n} d_n(t) < M\ .$$

Those of a right-window cutoff are optimal if for any $c > 0$:

$$\limsup_{n\to\infty} \sup_{t>t_n+cw_n} d_n(t) > 0\ .$$

This corresponds to strong optimality in the sense of [4, Definition 2.2]. Of course, if a profile cutoff holds, then the left- and right-window cutoffs hold at the same location and width, which are optimal for both. Examples will be given in section 3.

Our main result relates the location and width of the left- and right-window cutoffs to the terms of a decomposition into exponentials of the functions $d_n(t)$. From now on, we assume $M = +\infty$: the distance is relative entropy, L^p for $p > 1$, etc. The result is expressed for a sequence of continuous time processes, it could be written in discrete time, at the expense of heavier notations.

Theorem 2.1. *Assume that for each n, there exist an increasing sequence of positive reals $(\rho_{i,n})$, and a sequence of non negative reals $(a_{i,n})$ with $a_{1,n} > 0$, such that:*

$$d_n(t) = \sum_{i=1}^{+\infty} a_{i,n}\, \mathrm{e}^{-\rho_{i,n} t}\ . \tag{1}$$

Denote by $A_{i,n}$ the cumulated sums of $(a_{i,n})$, truncated to values no smaller than 1.

$$A_{i,n} = \max\{1, a_{1,n} + \cdots + a_{i,n}\}\ .$$

For each n, define:

$$t_n = \sup_i \frac{\log(A_{i,n})}{\rho_{i,n}}\ , \tag{2}$$

$$w_n = \frac{1}{\rho_{1,n}}\ , \tag{3}$$

$$r_n = w_n \left(\log(\rho_{1,n}t_n) - \log(\log(\rho_{1,n}t_n))\right)\ . \tag{4}$$

Assume that:

1. *for n large enough,*
$$0 < t_n < +\infty\ , \tag{5}$$

2. $$\lim_{n\to\infty} \rho_{1,n} t_n = +\infty\ , \tag{6}$$

3. *there exists a positive real α such that for n large enough, and for all $i \geqslant 2$,*
$$a_{i,n} \leqslant \alpha A_{i-1,n}\ . \tag{7}$$

Then (X_n) has a left-window cutoff at (t_n, w_n), a right-window cutoff at $(t_n + r_n, w_n)$. More precisely:

$$\forall c < 0 \ , \quad \liminf_{n\to\infty} d_n(t_n + cw_n) \geqslant \mathrm{e}^{-c} \ , \tag{8}$$

$$\forall c > 0 \ , \quad \limsup_{n\to\infty} d_n(t_n + r_n + cw_n) \leqslant \mathrm{e}^{-c} \ . \tag{9}$$

Conditions (5) and (7) are technical. Condition (6) is known as Peres criterion: Chen and Saloff-Coste [4] have proved that it implies cutoff for L^p distances with $p > 1$, and given a counterexample for the L^1 distance. A consequence is that $w_n = o(t_n)$ as requested by Definition 2.1, and more precisely that $w_n = o(r_n)$ and $r_n = o(t_n)$.

A decomposition into exponentials of the distance to equilibrium such as (1) holds for many processes: functions of finite state space Markov chains, functions of exponentially ergodic Markov processes, etc. Assuming that the decomposition only has non-negative terms is a stronger requirement: see [5, section 4]. It implies that $d_n(t)$ is a decreasing function of t. We do not view it as a limitation. Indeed, if (1) has negative terms, it can be decomposed as $d_n(t) = d_n^+(t) - d_n^-(t)$, with:

$$d_n^+(t) = \sum_{i=1}^{+\infty} \max\{a_{i,n}, 0\}\, \mathrm{e}^{-\rho_{i,n} t} \quad \text{and} \quad d_n^-(t) = -\sum_{i=1}^{+\infty} \min\{a_{i,n}, 0\}\, \mathrm{e}^{-\rho_{i,n} t} \ .$$

Assume that Theorem 2.1 applies to both $d_n^+(t)$ and $d_n^-(t)$, leading to left-window cutoffs at (t_n^+, w_n^+) and (t_n^-, w_n^-), right-window cutoffs at $(t_n^+ + r_n^+, w_n^+)$ and $(t_n^- + r_n^-, w_n^-)$. Since $d_n(t)$ is nonnegative, $t_n^- \leqslant t_n^+$, $t_n^- + r_n^- \leqslant t_n^+ + r_n^+$, and $w_n^- < w_n^+$. The sequence (X_n) has a right-window cutoff, and (9) holds for d_n with $(t_n + r_n, w_n) = (t_n^+ + r_n^+, w_n^+)$. Moreover, if $t_n^- + r_n^- = o(t_n^+)$ then the sequence (X_n) has a left-window cutoff, and (8) holds for d_n with $(t_n, w_n) = (t_n^+, w_n^+)$.

Theorem 3.8 in [5] contains a less tight assertion: it describes a (t_n, r_n)-cutoff, which can be deduced from Theorem 2.1 above. However, it hides the fact that when there is a (two-sided) window cutoff, the optimal width is no larger than w_n thus strictly smaller than r_n. The latter quantity is a correction bound on the location rather than a width: the optimal location may be anywhere between t_n and $t_n + r_n$.

In the next section, sequences of processes having a profile cutoff at (t_n, w_n) or $(t_n + r_n, w_n)$, with profile $F(c) = \mathrm{e}^{-c}$ will be constructed, thus proving that (8) and (9) are tight.

3 Examples

Several examples from the existing literature could be written as particular cases of Theorem 2.1: reversible Markov chains for the L^2 distance [17, 5], n-tuples of independent processes for the relative entropy distance [2], random walks on sums or products of graphs [19], samples of Ornstein-Uhlenbeck processes [11]. The objective of this section is not an extensive review of possible applications, but rather the explicit construction of some sequences illustrating the tightness of (8) and (9), and the possible locations of window cutoffs. We shall use here the relative entropy distance, also called Kullback-Leibler divergence: if μ and ν are two probability measures with densities f and g with respect to λ, then:

$$d(\mu, \nu) = \int_{S_\mu} f \log(f/g)\, \mathrm{d}\lambda \ ,$$

where S_μ denotes the support of μ. The main advantage of choosing that distance is its simplicity for dealing with tensor products:

$$d(\mu_1 \otimes \mu_2, \nu_1 \otimes \nu_2) = d(\mu_1, \nu_1) + d(\mu_2, \nu_2) \ .$$

Let a and ρ be two positive reals. Our building block will be a one-dimensional Ornstein-Uhlenbeck process, denoted by $X_{a,\rho}$ (see Lachaud [11] on cutoff for samples of Ornstein-Uhlenbeck processes). The process $X_{a,\rho}$ is a solution of the equation:

$$\mathrm{d}X(t) = -\frac{\rho}{2}X(t)\,\mathrm{d}t + \sqrt{\rho}\,\mathrm{d}W(t)\;,$$

where W is the standard Brownian motion. The distribution of $X_{a,\rho}(0)$ is normal with expectation $\sqrt{2a}$ and variance 1. It can be easily checked that the distribution of $X_{a,\rho}(t)$ is normal with expectation $\sqrt{2a}\,\mathrm{e}^{-\rho t/2}$ and variance 1. Therefore the (relative entropy) distance to equilibrium is:

$$d(t) = a\,\mathrm{e}^{-\rho t}\;.$$

Consider now two sequences (a_n) and (ρ_n) of positive reals, and assume that (a_n) tends to infinity. Theorem 2.1 applies to the sequence of processes (X_{a_n,ρ_n}) with $a_{1,n} = a_n$, $\rho_{1,n} = \rho_n$, and $a_{i,n} = 0$ for $i > 1$. The location and width are:

$$t_n = \frac{\log(a_n)}{\rho_n} \quad \text{and} \quad w_n = \frac{1}{\rho_n}\;.$$

The sequence has a profile cutoff at (t_n, w_n) with profile $F(c) = \mathrm{e}^{-c}$. Indeed:

$$d_n(t_n + cw_n) = a_n\mathrm{e}^{-(\rho_n t_n + c)} = \mathrm{e}^{-c}\;.$$

Hence (8) is tight. For $\rho_n \equiv \rho$, $X_{a_n,\rho}$ is a Markov process with a fixed semigroup, and an increasingly remote starting point: cutoff for such sequences were studied in [14].

Using tuples of independent Ornstein-Uhlenbeck processes, one can construct sequences X_n for which the distance to equilibrium is any finite sum of exponentials. Let m_n be an integer. For $i = 1, \ldots, m_n$, let $a_{i,n}$ and $\rho_{i,n}$ be two positive reals. Define the process X_n as:

$$X_n = \left(X_{a_{1,n},\rho_{1,n}}, \ldots, X_{a_{m_n,n},\rho_{m_n,n}}\right)\;,$$

where the coordinates are independent, each being an Ornstein-Uhlenbeck process as defined above. The distance to equilibrium of X_n is:

$$d_n(t) = \sum_{i=1}^{m_n} a_{i,n}\,\mathrm{e}^{-\rho_{i,n}t}\;. \tag{10}$$

Let n be an integer larger than 1. Let β_n be a real such that $0 \leqslant \beta_n \leqslant 1$. Define:

$$a_{1,n} = \mathrm{e}^n\;, \quad \rho_{1,n} = \frac{n}{1 + \frac{\beta_n}{n}\log\left(\frac{n}{\log(n)}\right)}\;, \tag{11}$$

and for $i = 2, \ldots, m_n = 9^n$,

$$a_{i,n} = \mathrm{e}^{-n}\;, \quad \rho_{i,n} = \log(\mathrm{e}^n + (i-1)\mathrm{e}^{-n})\;. \tag{12}$$

The following notation is introduced for clarity:

$$\ell_n = \log\left(\frac{n}{\log(n)}\right)\;.$$

Using (2), (3), and (4), one gets:

$$t_n = 1 + \frac{\ell_n\beta_n}{n} = \frac{n}{\rho_{1,n}}\;, \quad w_n = \frac{t_n}{n}\;, \quad r_n = \frac{t_n\ell_n}{n} = \ell_n w_n\;. \tag{13}$$

Lemma 3.1. *Let d_n be defined by (10), with $a_{i,n}$ and $\rho_{i,n}$ given by (11) and (12).*

Assume the following limit (possibly equal to $+\infty$) exists:

$$\gamma = \lim_{n\to\infty} (1-\beta_n)\ell_n \; . \tag{14}$$

Then:

$$\forall c \in \mathbb{R} \; , \quad \lim_{n\to\infty} d_n \left(t_n + (1-\beta_n) r_n + c w_n\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma}) \; . \tag{15}$$

A few particular cases are listed below. They illustrate the variety of possible behaviors.

- $\beta \equiv 1$: a cutoff with profile $2\mathrm{e}^{-c}$ occurs at (t_n, w_n).
- $\beta_n \equiv \beta \in [0,1)$: a cutoff with profile e^{-c} occurs at $(t_n + (1-\beta)r_n, w_n)$. For $\beta = 0$, this proves that (9) is tight.
- $\beta_n = (1+(-1)^n)/2$: a left-window cutoff occurs at (t_n, w_n), a right-window cutoff at $(t_n + r_n, w_n)$. The locations and width are optimal.
- $\beta_n = 1 - \gamma/\ell_n$, with $\gamma > 0$: a cutoff with profile $\mathrm{e}^{-c}(1+\mathrm{e}^{\gamma})$ occurs at (t_n, w_n).
- $\beta_n = 1 - (2+(-1)^n)/\ell_n$: a (t_n, w_n)-cutoff occurs, t_n and w_n are optimal. Yet no value of c is such that $d_n(t_n + c w_n)$ converges: there is no profile.

Proof. The main step is the following limit.

$$\lim_{n\to\infty} d_n \left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma}) \; . \tag{16}$$

In the sum defining d_n, let us isolate the first term: $d_n\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right) = D_1 + D_2$, with

$$D_1 = a_{1,n} \exp\left(-\rho_{1,n}\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)\right) \text{ and } D_2 = \sum_{i=2}^{m_n} a_{i,n} \exp\left(-\rho_{i,n}\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)\right) \; .$$

The first term is:

$$D_1 = \exp\left(-\frac{(1-\beta_n)\ell_n + c}{t_n}\right) \; .$$

Its limit is $\mathrm{e}^{-(\gamma+c)}$ because $(1-\beta_n)\ell_n$ tends to γ and t_n tends to 1. The second term is:

$$D_2 = \sum_{i=2}^{+\infty} \mathrm{e}^{-n} \left(\mathrm{e}^{n} + (i-1)\mathrm{e}^{-n}\right)^{-\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)} \; .$$

Thus D_2 is a Riemann sum for the decreasing function $x \mapsto x^{-\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)}$. Therefore,

$$\int_{\mathrm{e}^{n}+\mathrm{e}^{-n}}^{\mathrm{e}^{n}+m_n\mathrm{e}^{-n}} x^{-\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)} \,\mathrm{d}x < D_2 < \int_{\mathrm{e}^{n}}^{\mathrm{e}^{n}+(m_n-1)\mathrm{e}^{-n}} x^{-\left(1 + \frac{\ell_n}{n} + \frac{c}{n}\right)} \,\mathrm{d}x \; . \tag{17}$$

Now:

$$\frac{(\mathrm{e}^{n})^{-\left(\frac{\ell_n}{n} + \frac{c}{n}\right)}}{\frac{\ell_n}{n} + \frac{c}{n}} = \mathrm{e}^{-c} \frac{\log(n)}{\ell_n + c} \; ,$$

which tends to e^{-c}. Moreover,

$$\frac{(\mathrm{e}^{n} + (m_n-1)\mathrm{e}^{-n})^{-\left(\frac{\ell_n}{n} + \frac{c}{n}\right)}}{\frac{\ell_n}{n} + \frac{c}{n}} \leqslant \frac{n}{\ell_n + c} \left(\frac{m_n^{1/n}}{\mathrm{e}}\right)^{-(\ell_n + c)} \; ,$$

which tends to 0 for $m_n = 9^n > \mathrm{e}^{2n}$. So the upper bound in (17) tends to e^{-c}. There remains to prove that the difference between the two integrals tends to 0. That difference is smaller than:

$$\int_{\mathrm{e}^n}^{\mathrm{e}^n+\mathrm{e}^{-n}} x^{-\left(1+\frac{\ell_n}{n}+\frac{c}{n}\right)}\,\mathrm{d}x = \left(\frac{(\mathrm{e}^n)^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}}{\frac{\ell_n}{n}+\frac{c}{n}}\right)\left(1-(1+\mathrm{e}^{-2n})^{-\left(\frac{\ell_n}{n}+\frac{c}{n}\right)}\right) .$$

We have seen that the first factor tends to e^{-c}. The second factor tends to 0, hence the result.

Let us now deduce (15) from (16). Using (13),

$$1+\frac{\ell_n}{n}+\frac{c}{n} = t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n} .$$

Hence:

$$\lim_{n\to\infty} d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right) = \mathrm{e}^{-c}(1+\mathrm{e}^{-\gamma}) . \tag{18}$$

Let us write:

$$t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n} = t_n + (1-\beta_n)r_n + cw_n - ((1-\beta_n)r_n + cw_n)\left(\frac{\ell_n\beta_n}{nt_n}\right) .$$

Therefore:

$$\begin{aligned}
0 &\leqslant d_n\left(t_n + (1-\beta_n)\frac{r_n}{t_n} + c\frac{w_n}{t_n}\right) - d_n\left(t_n + (1-\beta_n)r_n + cw_n\right)\\
&\leqslant \left(\exp\left(\rho_{1,n}\left(((1-\beta_n)r_n + cw_n)\frac{\ell_n\beta_n}{n}\right) - 1\right)\right) d_n\left(t_n + (1-\beta_n)r_n + cw_n\right)\\
&= \left(\exp\left(\frac{\ell_n^2(1-\beta_n)\beta_n + c\ell_n\beta_n}{nt_n}\right) - 1\right) d_n\left(t_n + (1-\beta_n)r_n + cw_n\right) .
\end{aligned}$$

Hence the difference tends to 0, since $\frac{\ell_n^2}{n}$ tends to 0. □

4 Proof of Theorem 2.1

Proofs of inequalities (8) and (9) are given below.

Proof of (8). Let c be a negative real. Fix ϵ such that $0 < \epsilon < -c$. Using (2), define i_n^* as:

$$i_n^* = \min\left\{ i\,,\ t_n - \epsilon w_n \leqslant \frac{\log(A_{i,n})}{\rho_{i,n}} \leqslant t_n \right\} . \tag{19}$$

From (6), $t_n + cw_n$ is positive for n large enough. Then:

$$\begin{aligned}
d_n(t_n + cw_n) &= \sum_{i=1}^{+\infty} a_{i,n}\exp(-\rho_{i,n}(t_n + cw_n))\\
&\geqslant \sum_{i=1}^{i_n^*} a_{i,n}\exp(-\rho_{i,n}(t_n + cw_n))\\
&\geqslant A_{i_n^*,n}\exp(-\rho_{i_n^*,n}(t_n + cw_n))\\
&\geqslant \exp((-\epsilon w_n - cw_n)\rho_{i_n^*})\\
&\geqslant \exp((-\epsilon w_n - cw_n)\rho_{1,n})\\
&= \mathrm{e}^{-c-\epsilon} .
\end{aligned}$$

Since the inequality holds for all $\epsilon > 0$, the result follows. □

Proof of (9). Let c be a positive real. Our goal is to prove the following inequality.

$$d_n(t_n + r_n + cw_n) \leqslant \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} \left(\frac{r_n + cw_n}{t_n} + \mathrm{e}^{C_n} \right) , \tag{20}$$

where C_n tends to 0 as n tends to infinity. Let us first check that (20) implies (9). Observe that $\frac{r_n+cw_n}{t_n}$ tends to 0. Using (3) and (4):

$$\mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} = \mathrm{e}^{-c} \frac{1}{1 - \frac{\log(\log(t_n\rho_{1,n}))+c}{\log(t_n\rho_{1,n})}} .$$

By (6) the right-hand side tends to e^{-c}, hence the result.

To prove (20), split the sum defining $d_n(t_n+r_n+cw_n)$ into two parts S_1 and S_2, with:

$$S_1 = \sum_{i=1}^{l} a_{i,n} \exp(-\rho_{i,n}(t_n+r_n+cw_n)) \quad \text{and} \quad S_2 = \sum_{i=l+1}^{+\infty} a_{i,n} \exp(-\rho_{i,n}(t_n+r_n+cw_n)) .$$

Using the fact that the $\rho_{i,n}$ are increasing,

$$S_1 \leqslant A_{l,n} \exp(-\rho_{1,n}(t_n + r_n + cw_n)) . \tag{21}$$

To bound S_2, the idea is the same as in the proof of (15). From (2), $\exp(-\rho_{i,n}t_n) \leqslant A_{i,n}^{-1}$. Therefore:

$$S_2 \leqslant \sum_{l+1}^{+\infty} a_{i,n} A_{i,n}^{-(1+(r_n+cw_n)/t_n)} . \tag{22}$$

The function $x \mapsto x^{-(1+(r_n+cw_n))/t_n}$ is decreasing, and its integral from l to $+\infty$ converges. The right-hand side of (22) is a Riemann sum for that integral. Therefore:

$$S_2 \leqslant \frac{t_n}{r_n + cw_n} A_{l,n}^{-(r_n+cw_n)/t_n} . \tag{23}$$

Consider first the particular case $t_n = \frac{\log(A_{1,n})}{\rho_{1,n}}$, or equivalently $A_{1,n} = \exp(t_n\rho_{1,n})$. Applying (21) and (23) for $l = 1$ yields:

$$d_n(t_n + r_n + cw_n) \leqslant \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} \left(\frac{r_n + cw_n}{t_n} + 1 \right) , \tag{24}$$

which is (20) for $C_n = 0$. Otherwise, $A_{1,n} < \exp(t_n\rho_{1,n})$. Let ϵ be such that $0 < \epsilon < (t_n\rho_{1,n} - \log(A_{1,n}))/w_n$. The index i_n^* defined by (19) is larger than 1. The set of integers l such that $A_{l,n} < \mathrm{e}^{\rho_{1,n}t_n}$, contains 1 and is bounded by i_n^*. Therefore, there exists $l_n > 1$ such that:

$$A_{l_n-1,n} < \mathrm{e}^{\rho_{1,n}t_n} \leqslant A_{l_n,n} . \tag{25}$$

Applying (21) and (23) to $l = l_n - 1$ yields:

$$\begin{aligned} d_n(t_n + r_n + cw_n) &\leqslant \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} + \frac{t_n}{r_n + cw_n} \exp\left(-\frac{r_n + cw_n}{t_n} \log A_{l_n-1,n} \right) \\ &= \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} + \frac{t_n}{r_n + cw_n} \exp\left(-(r_n + cw_n)\rho_{1,n} \frac{\log A_{l_n-1,n}}{\rho_{1,n}t_n} \right) \\ &= \mathrm{e}^{-(r_n+cw_n)\rho_{1,n}} \frac{t_n}{r_n + cw_n} \left(\frac{r_n + cw_n}{t_n} + \mathrm{e}^{C_n} \right) . \end{aligned} \tag{26}$$

with

$$C_n = (r_n + cw_n)\rho_{1,n} \left(1 - \frac{\log A_{l_n-1,n}}{\rho_{1,n}t_n} \right) . \tag{27}$$

We must prove that C_n tends to 0. By (3) and (4):

$$(r_n + cw_n)\rho_{1,n} = \log(\rho_{1,n}t_n) - \log\log(\rho_{1,n}t_n) + c \; . \tag{28}$$

From (25):

$$0 < 1 - \frac{\log(A_{l_n-1,n})}{\rho_{1,n}t_n} \leqslant \frac{1}{\rho_{1,n}t_n} \log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) \; . \tag{29}$$

Plugging (28) and (29) into (27), for n large enough:

$$0 < C_n \leqslant \left(\frac{\log(\rho_{1,n}t_n) - \log\log(\rho_{1,n}t_n) + c}{\rho_{1,n}t_n}\right) \log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) \; .$$

By (6), the first factor of the right-hand side tends to 0. Moreover, condition (7) entails that for n large enough:

$$\log\left(1 + \frac{a_{l_n,n}}{A_{l_n-1,n}}\right) < \log(1+\alpha) \; .$$

Hence the result. □

References

[1] D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348.

[2] J. Barrera, B. Lachaud, and B. Ycart, *Cutoff for n-tuples of exponentially converging process*, Stochastic Process. Appl. **116** (2006), no. 10, 1433–1446.

[3] D. Bayer and P. Diaconis, *Trailing the dovetail shuffle to its lair*, Ann. Appl. Probab. **2** (1992), no. 2, 294–313.

[4] G. Y. Chen and L. Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, Electron. J. Probab. **13** (2008), no. 3, 26–78.

[5] ———, *The L^2-cutoff for reversible Markov processes.*, J. Funct. Anal. **258** (2010), no. 7, 2246–2315.

[6] P. Diaconis, R. Graham, and J. Morrison, *Asymptotic analysis of a random walk on a hypercube with many dimensions*, Random Struct. Algor. **1** (1990), no. 1, 51–72.

[7] P. Diaconis and L. Saloff-Coste, *Separation cut-offs for birth and death chains*, Ann. Appl. Probab. **16** (2006), no. 4, 2098–2122.

[8] P. Diaconis and M. Shahshahani, *Time to reach stationarity in the Bernoulli-Laplace diffusion model*, SIAM J. Math. Anal. **18** (1987), no. 1, 208–218.

[9] A. Diédhiou and P. Ngom, *Cutoff time based on generalized divergence measure*, Statist. Probab. Lett. **79** (2009), no. 10, 1343–1350.

[10] J. Ding, E. Lubetzky, and Y. Peres, *Total-variation cutoff in birth-and-death chains*, Probab. Theory Rel. Fields **146** (2010), no. 1-2, 61–85.

[11] B. Lachaud, *Cutoff and hitting times for a sample of Ornstein-Uhlenbeck processes and its average*, J. Appl. Probab. **42** (2005), no. 4, 1069–1080.

[12] B. Lachaud and B. Ycart, *Convergence times for parallel Markov chains.*, Positive systems. Proceedings of the second multidisciplinary international symposium on positive systems: Theory and applications (POSTA 06), Grenoble, France, August 30 – September 1, 2006, Springer, Berlin, 2006, pp. 169–176.

[13] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2006.

[14] S. Martínez and B. Ycart, *Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space*, Adv. Appl. Probab. **33** (2001), no. 1, 188–205.

[15] J. Miller and Y. Peres, *Uniformity of the uncovered set of random walk and cutoff for lamplighter chains*, Ann. Probab. **40** (2012), no. 2, 535–577.

[16] L. Saloff-Coste, *Random walks on finite groups*, Probability on discrete structures, Encyclopaedia Math. Sci., vol. 110, Springer, Berlin, 2004, pp. 263–346.

[17] B. Ycart, *Cutoff for samples of Markov chains.*, ESAIM: P&S **3** (1999), 89–106.

[18] ———, *Stopping tests for Markov chain Monte-Carlo methods.*, Methodol. Comput. Appl. Probab. **2** (2000), no. 1, 23–36.

[19] ———, *Cutoff for large sums of graphs*, Ann. Inst. Fourier **57** (2007), no. 7, 2197–2208.

Acknowledgements: J. Barrera was partially supported by grants Anillo ACT88, Fondecyt n°1100618, and Basal project CMM (Universidad de Chile). B. Ycart was supported by Laboratoire d'Excellence TOUCAN (Toulouse Cancer).